DOCUMENT RESUME

ED 139 255 FL 008 547

AUTHOR Irvine, David J.
TITLE Evaluation Methodology for a French Language Immersion Program.
PUB DATE Mar 77
NOTE 9p.; Paper presented at the Evaluation Workshop on the French Language Immersion Program (State University College, Plattsburgh, New York, March 17-18, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Bilingual Education; Elementary Education; *Evaluation Methods; *French; Immersion Programs; *Language Instruction; *Language Programs; Language Skills; Measurement Techniques; Program Effectiveness; *Program Evaluation; Research Design; *Second Language Learning
IDENTIFIERS New York (Plattsburgh)

ABSTRACT

The Plattsburgh French Language Immersion Program (FLIP) is designed to enrich the education of majority-group children. In discussing evaluation methodology for FLIP, the general problems of evaluating a program are discussed, as well as reactions to details of the design developed by the research group for FLIP. Evaluation is distinguished from research as the determination of how well a particular program is working and what processes account for its effectiveness. The research design for FLIP is criticized in terms of the number of measures called for, the small size of the experimental and control groups which possibly influences results, the form in which hypotheses are stated, the way in which academic achievement is measured, and the fact that a number of variables are not controlled for. Questions raised concern which theory of child development will be the basis for research, how FLIP fits into such a scheme and the kinds of questions that need to be answered about FLIP, what kinds of data are needed to do justice both to underlying philosophical and theoretical questions and to FLIP, and how best to collect and analyze such data. (CLK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139255

FL

EVALUATION METHODOLOGY FOR A FRENCH LANGUAGE IMMERSION PROGRAM

David J. Irvine
Coordinator
Prekindergarten Evaluation Unit
State Education Department
Albany, New York 12234

A paper presented at the Evaluation Workshop on the French Language Immersion Program, State University College, Plattsburgh, New York, March 17-18, 1977.

FL008547

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATION METHODOLOGY FOR A FRENCH LANGUAGE IMMERSION PROGRAM

The Plattsburgh French Language Immersion Program (FLIP) is an interesting innovation designed to enrich the education of majority-group children. In discussing evaluation methodology for FLIP, I will break my comments into three sections: 1) a framework, to let you know how I view the problems of evaluating a program; 2) reactions to details of the design developed by the research group for FLIP; and 3) some underlying questions.

A Framework for Evaluation

As a beginning, it might be useful to distinguish between research and evaluation. Research is aimed toward uncovering new knowledge. Evaluation, on the other hand, is designed to determine how well a particular program (or method, or set of materials) is working and what processes account for its effectiveness. There is great overlap between the two, since many projects are designed to provide both new knowledge and an assessment of the effectiveness of the program being studied. And evaluation of a program may contribute, as a by-product, to the pool of information about programs of the same kind. In addition, the two areas use many of the same techniques and are staffed by people with similar competencies. FLIP, then, incorporates some elements of research and some of evaluation. Rather than shift back and forth between the two terms, I will talk about evaluation, with the understanding that much of what is said could apply as well to research.

Evaluation is a human process which is going on all of the time. It is a common part of everyday living. It can be called a human process, also, because it goes on inside of human beings. Tests, statistics, and computers do not evaluate. They are merely tools--exceedingly useful tools--which are used to support the human process of evaluation. Neither a test, a statistic, nor a computer can say whether a program has been effective. But they can be used by a human being to generate evidence which he or she can use in judging whether the program has been effective.

It is important that we formalize the evaluations which we are carrying on and that we have standard language and conventions for describing those evaluations. We do this to make our objectives explicit; to force us to articulate what we know and what we want to find out; to make it possible for us to repeat a program; to communicate it to others so they can duplicate it; and, not least important, to communicate our results to others in order to induce them to support the program and similar programs.

It would be wonderful if a college dean or president or legislator would say to a program director, "I know that you're a fine fellow, so if you think that program should be funded, so be it." Wonderful, but unlikely. Even our most supportive leaders are more likely to say, "I know that you're a fine fellow, so I would appreciate it if you would share with me the evidence that convinces you that your program should be funded." At that point, a momentum is begun toward more formal and systematic evaluation.

It is also a fact of life that the more remote the person whom we are trying to persuade, the more we must communicate through mutually understood language and conventions, including such things as tests and statistics.

Formal evaluation involves systematically reducing the number of possible explanations for a given phenomenon. We want to eliminate, insofar as we are able, the possibility that children's status at the end of a program was due to their high status at the beginning, or that they had advantages over children in other programs or children not in the program, or that things outside the program actually may account for the results. In short, to evaluate the impact of a program, we systematically eliminate other possible explanations of the participants' performance until we can say with a specified level of confidence that the results we obtained were not due to chance or to unrelated influences. Then we are justified in making inferences about the impact of the program.

Reactions to the FLIP Research Design

The research team for the French Language Immersion Program is grappling with a number of knotty problems. Many of the problems stem from the program and the context within which it operates rather than originating in the research design.

Several specific points might be made about det ils of the research design (Singson and Hornby, undated). First, a large number of measures are called for. With the number of measures used, they can hardly be unobtrusive. It may be desirable to try to trim down the number of measures. A matrix of simple correlations between measures might provide insight into possible overlap. To carry the procedure further, a factor analysis might be carried out of all the measures, including their subtests. This would show the underlying structures which are being measured and might make it possible to reduce the number of measures or reorganize the measures within the factor structure. A more clearly

articulated theoretical position in the research design might provide a justification for the measures selected by helping to identify areas to be studied and areas not relevant to the study.

Although no statistically significant differences were found between the experimental group (children in the FLIP class) and the control group, the small sizes of the two groups would require quite large differences on most of the measures to be significant. For example, a difference of approximately six IQ points was found between the two groups on the Wechsler Intelligence Scale for Children. This difference was found to be nonsignificant, presumably because of the small number in the groups. Yet it appears to be large enough to raise questions about the comparability of the two groups. In general, the FLIP group appears to be favored somewhat on a number of cognitive measures while the control group is favored on the measure of self concept.

The children in the experimental and control groups are stratified by age and sex. This results in extremely small numbers in some of the cells. Since age is a continuous variable, it appears that a better strategy would be to use it as a covariate in analyses of covariance. Multivariate analysis of covariance would permit such handling of several variables, might reduce the seriousness of the problem of the small numbers of subjects, and would permit study of interactions between such variables as age, sex, pretest scores, and children's characteristics.

The small number of children can be both an opportunity and a bother. With small numbers it may be possible to investigate in depth a variety of questions often overlooked in large-scale research projects. The small numbers are troublesome when statistical operations are attempted, and when you want to generalize your results to children in general.

A number of interesting process variables could be investigated. Since process evaluation was dealt with by another speaker at this conference (Bruck, 1977), elaboration in the presentation would be redundant. However, studying such variables as classroom procedures, structure, and interactions might be extremely fruitful.

The research plan presents five specific research hypotheses. Two are stated in affirmative terms. For consistency, they could all be stated in the same form, preferably in the null form. Use of the null form may seem to be playing semantic games. However, it serves a useful purpose by encouraging conservative inferences from the data. For example, saying, "I have no evidence that the two groups are different" (nonsignificant results based on a null hypothesis) is quite different from saying, "I have evidence that the two groups are alike" (nonsignificant results based on an affirmative hypothesis).

Academic achievement is to be measured by the Wide Range Achievement Test. The research plan states that "evaluation will be directed at grade level scores" in the three areas measured by the test. If this means that Grade Equivalent Scores will be used as data for analysis, problems may be encountered. Grade Equivalents have severe psychometric weaknesses. A preferred procedure would be to analyze raw scores. Later, findings might be converted to Grade Equivalents, if that aids in interpretation, but that would come after the analysis, avoiding problems of manipulating statistically Grade Equivalents.

Finally, the most troublesome point of all: In the research design, several variables are not controlled. It should be pointed out that this is primarily a result of the conditions under which the program operates. Nevertheless, it will be extremely difficult to draw unequivocal inferences from the results of the study. Uncontrolled variables include:

1. The treatment, instruction in French versus instruction in English.

2. The treatment, French Immersion as a technique vs. non-immersion.

3. Highly structured instruction (in the FLIP classroom) versus openness in instruction (in the control group).

4. Teacher differences.

If children in the FLIP program do better on a given criterion than the control children, will it be because they received instruction in French, because they received instruction through the specific technique called French Immersion, because they were taught in a highly structured classroom, or because of the particular teacher they had? It will be difficult to know.

This problem might be overcome as additional FLIP classes and teachers are added. A second FLIP class might be taught in a less teacher-centered manner. A control-group class might be taught in a more highly teacher-centered manner. Process variables could be studied in all of the classrooms, experimental and control, in order to determine their impact on learning in the different settings.

Some Underlying Questions

The FLIP research design suggests a number of questions which have to do with the educational needs of children and how they might be met by FLIP. Merely raising such questions is to be applauded, for they have much to do with the focus of both FLIP and research regarding FLIP.

A sequence of questions might be framed to follow those raised in the research design. Upon what kind of theory about the way children develop are we going to operate? How does FLIP fit into such a scheme? What kinds of questions do we want to answer about FLIP? What kinds of data are needed to do justice

both to the underlying philosophical and theoretical questions and to FLIP? What are the best ways of collecting and analyzing such data? These questions have been answered in part by the decisions made in designing the evaluation.

Such questions impact on the more specific question of "Is FLIP effective? The real effectiveness of the program will ultimately rest on value judgments made about the total learnings of the children in the context of the purposes of education. The statistical evidence which will be generated in the course of the evaluation cannot shape the context, but the context will shape the precise meaning of the statistical evidence.

REFERENCES

Bruck, Maggie. "Process vs. Product Research," a paper presented at the Evaluation Workshop on the French Language Immersion Program, State University College, Plattsburgh, New York, March 17-18, 1977.

Singson, Maribel R., and Peter A. Hornby. Development of Research Goals for the Plattsburgh French Language Immersion Program. Plattsburgh, New York: Margaret M. Sibley School for Educational Research and Demonstration, State University College. (Undated.)